



UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

January 11, 2000

DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2

MEMORANDUM FOR Howard Hogan
 Chief, Decennial Statistical Studies Division

From: Danny R. Childers *DRC*
 Debbie Fenstermaker *DF*
 Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation: Overview of the Design

Attached is a summary of the documentation of the design of the Census 2000 Accuracy and Coverage Evaluation. This is intended as an overview of the major steps of the operational and sampling design. For more details, see "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1. If there are any questions, contact Danny R. Childers (301-457-4184) or Debbie Fenstermaker (301-457-4195).

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
 A. C. E. Implementation Team
 Statistical Design Team Leaders
 Chron

The Design of the Census 2000 Accuracy and Coverage Evaluation

1.0 Introduction

The Census Bureau will conduct an Accuracy and Coverage Evaluation (A. C. E.) to measure the overall and differential coverage of the U.S. population in Census 2000. The A. C. E. will also provide base population figures for other Census Bureau programs, such as the Census Bureau's intercensal population estimates, American Community Survey, and other demographic surveys. Under the traditional census plan, the A. C. E. will not be used to adjust the census figures for reapportionment purposes.

The "Population Sample" or "P-sample" and "Enumeration Sample" or "E-sample" have traditionally defined the samples for dual system estimation. The P-sample consists of people enumerated independent of the census. The E-sample consists of people enumerated in the census. After matching and reconciliation, the P-sample yields an estimate of the population missed in the census while the E-sample yields an estimate of the correctly enumerated people in the census. Combining these two components yields an A. C. E. estimate of census coverage.

The major steps of the A. C. E. include various stages of sampling, matching, and field activities. See the Attachment for a flowchart of A. C. E. activities. Note that the sampling and operational activities are intermixed. In particular, most sampling activities occur before housing unit matching. After the initial A. C. E. sample is selected, the housing units within the sample block clusters are listed. The sample is reduced to provide the A. C. E. cluster design. Then, the A. C. E. housing units are matched to the census inventory of housing units. After reconciling the nonmatches, a list of A. C. E. housing units that are confirmed to have existed within the block clusters is prepared and large block subsampling is conducted. The housing units selected in sample define the P-sample of housing units. Person interviews are conducted for these P-sample housing units.

This memorandum outlines the major steps in the sampling and operational activities for A. C. E. For more details about the design of the A. C. E., see "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1. Additional discussions of the missing data and dual system estimation activities can be found in "Accuracy and Coverage Evaluation Survey: Missing Data Procedures", DSSD Census 2000 Procedures and Operations Memorandum Series Q and in "Accuracy and Coverage Evaluation Survey - Dual System Estimation", DSSD Census 2000 Procedures and Operations Memorandum Series Q.

2.0 The Initial Housing Unit Phase

2.1 The Listing Sample

As a result of the January 1999 Supreme Court ruling against the use of sampling for apportionment, the Census Bureau redesigned the Integrated Coverage Measurement (ICM) Survey as an A. C. E. The ICM was planned as a 750,000 housing unit sample while the A. C. E. sample is planned to be approximately 300,000 housing units. By the time of the Supreme Court decision, earlier commitments had become operationalized based on the ICM sample design, and consequently, the A. C. E. sample design had to be derived from the ICM design. Therefore, the entire ICM sample was selected as originally planned and then reduced through various stages to yield the target housing unit sample size.

Under the ICM sampling plan, key features of the A. C. E. listing sample selection include:

- roughly equal sample sizes for most states except the most populous
- a separate sample for American Indian Reservations
- roughly proportional allocation of sample within states
- a separate sample of small block clusters
- an oversample of large block clusters

The A. C. E. primary sampling unit is the block cluster. A block cluster is a single census collection block or group of geographically contiguous census collection blocks. Using housing unit counts from an early census address list, block clusters were stratified by size: small (0 to 2 housing units), medium (3 to 79 housing units) and large (80 or more housing units). In states with a sufficient number of American Indians living on reservations based on the 1990 census count, a separate sampling stratum was formed of American Indian Reservation block clusters. Within each sampling stratum, a systematic sample of block clusters was selected with equal probability.

The initial listing sample was selected in the second quarter of 1999. This stage of sampling yielded 29,136 block clusters and roughly 2 million housing units to be listed in the 50 states and the District of Columbia. For Puerto Rico, there were 559 block clusters in the listing sample and roughly 50,000 housing units to be listed.

2.2 Independent Listing

An independent listing of the addresses of all the housing units in the A. C. E. sample clusters is conducted before census day. The A. C. E. housing units are the housing units recorded in the Independent Listing Books (ILB). Besides listing each housing unit in the cluster, the listers will inquire about housing units present at each special place and commercial structure to obtain any additional housing units.

After the listing books are received in the National Processing Center (NPC) in Jeffersonville, Indiana, they are checked-in and keyed. The quality assurance is 100 percent for the keying of these listing books. Keying rejects are reviewed clerically to correct errors before the matching begins. A data file of the A. C. E. housing units is created for matching. The A. C. E. maps will be scanned at the NPC for use in the housing unit matching.

2.3 Medium and Large Block Cluster Reduction

The A. C. E. cluster reduction implementation for medium and large block clusters is scheduled for December 1999 through mid-January 2000, following two major operations: the completion of the A. C. E. independent listing operation and an update of the census address list. The resulting national sample allocation will be roughly proportional to state population with some differential sampling within states.

One component of the A. C. E. cluster reduction design is stratifying the A. C. E. listing clusters based on the relationship of current housing unit counts from the A. C. E. independent listing and the updated census address list. Clusters will be differentially sampled in order to reduce the variance contribution due to inconsistencies between the census and the independent list. Clusters with significant differences between the counts are likely to have high erroneous enumerations and high nonmatch rates. The objective of differentially sampling these types of clusters is to reduce the magnitude of the weights associated with clusters having potentially high variance contributions.

Another component of the A. C. E. cluster reduction design is differentially sampling clusters based on the estimated demographic composition of the cluster. Clusters with a high proportion of a minority or Hispanic group are classified into a minority stratum. The objective of differentially sampling these types of clusters is to increase the sample size and improve the reliability of the A. C. E. population estimates for these subgroups.

The A. C. E. cluster reduction occurs before the start of housing unit matching in order to reduce the volume of clusters going into that operation. Following this reduction there are expected to be roughly 15,500 block clusters in the A. C. E. reduced sample for the 50 states and the District of Columbia. The number of block clusters in Puerto Rico is unchanged at 559 block clusters.

2.4 Small Block Cluster Reduction

The small block cluster reduction occurs in late January through early February 2000 following the completion of the keying of the independent listing books. Small clusters are expected to have between zero and two housing units based on the early census address list. Conducting interviews and follow-up operations in small block clusters is not cost-effective compared to large clusters. Therefore, to allocate A. C. E. resources more efficiently, a subsample of these small clusters will be selected for the A. C. E. interviewing sample.

The small block cluster reduction has three goals. The first goal is to avoid having small cluster weights that are extremely high compared to other cluster weights in the sample. The second goal is to have lower weights on small clusters where the number of housing units is different from the early census address list. These two goals attempt to reduce the contribution of small clusters to the variance of the dual system estimates. The third goal is to ensure that the field staff can efficiently manage the resulting workloads.

Using the keyed independent listing counts and the currently available census counts, the small block clusters within each state are stratified by size. Then, a systematic subsample is selected from each stratum with equal probability. All small block clusters which have 10 or more housing units, and are either List Enumerate or on an American Indian Reservation or other type of American Indian Country are kept in the A. C. E. sample.

Approximately 1500 of the 5000 small block clusters are expected to be retained in the A. C. E. sample. This is the last stage of cluster reduction; therefore, there is expected to be approximately 12,500 block clusters in the A. C. E. sample for the 50 states and the District of Columbia. For Puerto Rico, there will be roughly 500 block clusters.

2.5 Initial Housing Unit Matching

The housing units included on the census address list in January 2000 in the block clusters after the block cluster reductions are obtained for the housing unit matching. The objective of housing unit matching is to link the A. C. E. and census housing units for automated subsampling in large blocks for both the P-sample and the E-sample. In addition, the cleaned-up list of P-sample addresses makes the person interviewing easier because all housing units have been confirmed to exist as housing units. The addresses for housing units in the A. C. E. and the census housing units in the A. C. E. block clusters are first computer matched. The computer matching is followed by an automated clerical review. There is also a clerical search, which is limited to the block cluster, for duplicate housing units during this phase of the matching. Possible duplicates in both the A. C. E. and the census are identified.

2.6 Housing Unit Follow-up

The cases coded as not matched, possibly matched, or possible duplicates after the matching are sent for follow-up interviews for all types of basic street address codes. Selected matched cases are also sent for additional information. Specifically, the cases identified for field follow-up are:

- The A. C. E. addresses with a before follow-up code of not matched are sent to the field to confirm if they were housing units within the block cluster.
- The census addresses with a before follow-up code of not matched are sent to the field to confirm if they were housing units within the sample block cluster.
- The possible matches are sent to the field to determine if the A. C. E. and census addresses refer to the same housing unit.

- Census housing units that are identified as possible duplicates are followed up to determine if the two census addresses refer to the same housing unit.
- A. C. E. housing units that are identified as possible duplicates are followed up to determine if the two A. C. E. addresses refer to the same housing unit.
- Matched housing units with a unit status code of under construction, future construction, unfit for habitation, vacant trailer site in a mobile home park, and other to determine the status of the housing unit at the time of the follow-up interview, since changes may have occurred since listing.

The questions on the follow-up form are not designed to be read to respondents but are used as a guide for an interviewer. The answer to one question may be the result of asking several other questions. The interviewer appropriately modifies the questions, when necessary, to the situation that is encountered in the field and records the appropriate answers on the follow-up form. Many questions can be answered by observation.

After the field follow-up, the completed forms are returned to the processing office. Using the information obtained during the field work, an After Follow-up Match Code is assigned for cases sent to the field.

2.7 Reduction Within Large Block Clusters

When block clusters are large, it is necessary to reduce the housing units within the cluster to obtain manageable field workloads for A. C. E. interviewing and person follow-up without having a big impact on reliability. Further, it is important to geographically overlap the P- and E-samples to reduce the E-sample person follow-up workload.

2.7.1 P-Sample

Following the completion of the housing unit matching and follow-up operations, the reduction of housing units within large block clusters is done. Any block cluster with 80 or more independently listed housing units after the matching and follow-up operation is eligible for this reduction. After the reduction within large block clusters is done, the interview sample size for the fifty states and the District of Columbia will be roughly 300,000 housing units. This stage of subsampling reduces the number of housing units in a cluster to be in the A. C. E. interview sample.

The reduction of housing units within a large block cluster is done by forming segments of adjacent housing units and selecting one or more segments for A. C. E. person interviewing. The segments have approximately equal numbers of housing units within a block cluster. Segments of housing units are used as the sampling unit in order to obtain compact interviewing workloads and to facilitate overlapping P- and E-samples to reduce E-sample person follow-up workloads.

2.7.2 E-Sample

The source of the E-sample is the Census Unedited File (CUF) which will be available in the fall of 2000. All census records in the A. C. E. block clusters are eligible to be in the E-sample. However, to reduce the person follow-up workload, a reduction of the census housing units is done when there are more than 80 census housing units in the block cluster.

To create overlapping P- and E-samples within a cluster when the census housing units are reduced, the results of the P-sample reduction results are mapped onto the census records on the CUF in the block cluster. An overlapping P- and E-sample is not necessary but is desired for cost effectiveness. This is possible because when there was a correspondence between an A. C. E. independently listed address and a census address during the initial housing unit matching, the census identification number was assigned to the A. C. E. unit. If there are a significant number of census addresses which do not correspond to an A. C. E. unit, then a second systematic housing unit reduction will be done to reduce the E-sample person follow-up workloads.

3.0 Person Interview

Three types of people are collected in the person interview: those who lived at the sample address at the time of the interview and on census day (i.e., nonmovers), those who have moved into the sample address since census day (i.e., inmovers), and those who lived at the sample address on census day but lived elsewhere at the time of the A. C. E. interview (i.e., outmovers). Census day residence status is established within the Computer Assisted Personal Interview (CAPI) instrument for the nonmovers and outmovers. This information is needed to prepare the dual system estimates.

The A. C. E. person interview is conducted using a CAPI instrument. In order to get an early start for the interviewing, a telephone interview will be conducted for households where the census questionnaire is data captured and included a telephone number. Both households with mail returns and enumerator filled questionnaires are eligible for telephone interviews. Housing units without house number and street name addresses and housing units in small multi-unit structures will be excluded from the telephone interviewing. Large multi-units are included in the telephone interviewing because they tend to have unique unit designations. Many small multi-unit structures and rural areas do not have addresses that allow the telephone interviewer to distinctly identify the address. We do not want to interview any addresses on the telephone without unique addresses. All remaining interviews after the telephone operation is completed will be conducted in person. However, some nonresponse conversion operation interviews and interviews in gated communities or secured buildings may be conducted by telephone.

The person interview is conducted only with a household member for the first three weeks of interviewing. If an interview with a household member is not obtained after three weeks, an interview with a nonhousehold member is attempted, called a proxy interview. The proxy

interviewing is allowed during the remainder of the interviewing period. During the last two weeks of interviewing a nonresponse conversion operation is attempted for the noninterviews using the interviewers considered to be the best available. This noninterview conversion will attempt to obtain an interview with a household member or a proxy respondent, but not a last resort interview¹. This noninterview conversion will reduce the noninterview rate.

After the names and characteristics are obtained, the resident status on census day is established. For nonmovers and outmovers, questions about mover status, group quarters, and other residences on census day establish the residence status.

4.0 The Person Phase

The P-sample people and the census people from the CUF are computer matched within cluster. The possible matches, P-sample nonmatches, and E-sample nonmatches are clerically reviewed using an automated computer match and review system. Additional matches and possible matches are identified by the clerical staff. Duplicates on both lists are also identified clerically. People with incomplete names are identified because they do not contain sufficient information for matching and follow-up. After the matching is completed, field follow-up is conducted for selected cases and the results of the field interview are coded in the matching database.

4.1 E-sample Insufficient Information for Matching and Follow-up

The census person records are reviewed to identify people with insufficient information for matching and follow-up. Only people with sufficient information for matching and follow-up are allowed to be processed in the matching and follow-up interviewing phases of the person matching. The three ways for a census record to be determined to have insufficient information are:

- The census people are not data defined.
- The census people are data defined, but computer coded as insufficient information for matching and follow-up.
- The census people are computer coded as sufficient information, but converted clerically to insufficient information for matching and follow-up.

The first type of census people who are not data defined are not included in the E-sample. Only data defined people are included in the E-sample. These data defined people create person records in the census. Sufficient information for matching and follow-up in the E-sample is complete name and two characteristics. See the design document referenced in Section 1.0 for a more detailed discussion of census data defined and insufficient information for matching and follow-up.

¹ A last resort interview is one obtained from a respondent who would not be classified as knowledgeable. This information may not include name and may contain minimal data.

4.2 P-sample Insufficient Information for Matching and Follow-up

The P-sample is reviewed by the computer software to identify people who have insufficient information for matching and follow-up. The P-sample rules for sufficient information for matching and follow-up are the same as the E-sample rules. Sufficient information for matching and follow-up in the P-sample is complete name and two characteristics. These cases identified by the computer will be suppressed from viewing by the clerical matchers to prevent errors in matching people who should not be converted to sufficient information for matching. The probability of matched will be imputed for the P-sample people coded as insufficient information for matching and follow-up. They are treated like other P-sample people with unresolved match status.

4.3 Person Matching

The people from A. C. E. housing units who will be initially matched to the E-sample and non E-sample census enumerations are:

- the nonmovers and outmovers identified as residents
- the people with unresolved residence status

The matching within the sample block clusters is done by the “computer matcher” followed by an automated clerical review. The computer matching will match the nonmovers and outmovers to the E-sample in subsampled blocks and then allow matching to the non E-sample census enumerations. These non E-sample enumerations are census people in housing units that were not included in the E-sample after the subsampling of census housing units. The goal of the matching is to produce the correct ratio of cases classified as omitted to those classified as included in the census.

Duplicates are identified in both the P-sample and E-sample people. A P-sample person duplicate is removed from the final P-sample. Whole households of P-sample duplicates are converted to noninterviews for Dual System Estimation. The A. C. E. interview was not a good interview when the whole household was duplicated. An E-sample person duplicate is an erroneous enumeration in the census.

4.4 Targeted Extended Search

The targeted extended search for 2000 A. C. E. is a two stage process. First, clusters are identified that will benefit most from expanding the search area to surrounding blocks because of geocoding error. Second, blocks within the cluster will be targeted for searching.

There are geocoding errors of exclusion and inclusion in the sample cluster. Geocoding errors of exclusion affect the P-sample nonmatch rate and geocoding errors of inclusion affect the E-sample erroneous enumeration rate. If the geocoding error omits the census housing unit from

the sample block cluster, the P-sample people and housing units will not be matched. Conversely, if the geocoding error includes the census housing unit in the sample block cluster, the E-sample people will be erroneously enumerated.

The clusters selected for targeted extended search for the 2000 Accuracy and Coverage Evaluation are:

- Clusters included with certainty:
 - Relisted clusters in A. C. E.
 - Five percent of the clusters with the most census geocoding errors and A. C. E. address nonmatches
 - Five percent of the clusters with the most weighted census geocoding errors and A. C. E. address nonmatches
- Clusters selected at random from the clusters with A. C. E. housing unit nonmatches or census housing units identified as geocoding errors.

Clusters without A. C. E. housing unit nonmatches and census geocoding errors are out-of-scope for the targeted extended search sampling. The initial housing unit matching results are used to identify the A. C. E. housing unit nonmatches and census housing unit geocoding errors. Any changes to the census inventory of housing units is not reflected in the housing unit matching used to identify targeted extended search clusters.

4.5 A. C. E. Person Follow-up

The person follow-up is conducted to gather additional information to accurately code the residence status of the nonmatched P-sample people and the enumeration status of the nonmatched E-sample people. The P-sample nonmatches do not match to the census. We want to make sure these P-sample nonmatches actually lived in the sample block cluster on census day. The P-sample nonmatch is sent for a follow-up interview when there is a possibility the residence status is not correct, such as partial household nonmatches, whole household nonmatches when the interview was obtained by a proxy interview, and when there is a conflicting household situation (i.e., Smith/Jones cases). The E-sample nonmatches are sent for a follow-up interview to determine if they were correctly or erroneously enumerated in the block cluster. We send possible matches for an interview to resolve their match status. There are also other cases sent to follow-up, such as matched people with unresolved residence status and other types of cases considered to have the potential for geographic errors in the P-sample.

The following table summarizes which P-sample nonmatches will be sent for a follow-up interview:

| Type of P-sample Nonmatch | Person Interview with a Proxy Respondent | Person Interview with a Household Member |
|--|---|---|
| Partial household nonmatch | Followed up | Followed up |
| Whole household nonmatch where the housing unit does not match | Followed up | Not followed up |
| Whole household nonmatch where the housing unit is matched to the census | Followed up | Not followed up |
| Whole household nonmatch with conflicting households | Followed up | Followed up |

In general, a partial household nonmatch is where there is at least one nonmatch and one matched P-sample person. A conflicting household is one where the address matches and both are occupied, but with different household members. For more information, see the detailed design document referenced in Section 1.0.

Also, we will add the following types of cases to follow-up to collect more information to verify P-sample housing unit geography in addition to asking the person nonmatch questions:

- All P-sample whole household nonmatches in relisted clusters, since they not included in the housing unit matching phase of the A. C. E.
- All P-sample whole household nonmatches in clusters with a high rate of P-sample person nonmatch. High has been defined as 45 percent.
- P-sample whole household nonmatches where the interviewer for the person interview changed the address for the P-sample housing unit. Information about the accuracy of the P-sample geography is obtained.
- Any P-sample whole household nonmatch identified by the Analyst as needing follow-up.

If the follow-up interview identifies a P-sample housing unit as a geocoding error, the people and housing units will be removed from the P-sample. The whole households of P-sample people incorrectly listed in the cluster will be coded as P-sample geocoding error.

The person follow-up is conducted using a paper questionnaire. The questionnaire is designed to gather information that may resolve matching and residence status problems. For example, a match between P-sample and E-sample people might be made if another piece of information is known. Also, information may be needed to confirm residence status on census day for matched or unmatched people. During the follow-up interview, interviewers will attempt to gather the information needed to code each person as a matched resident/non-resident or a nonmatched resident/non-resident of the block cluster on census day. The questionnaire places more emphasis on obtaining a good respondent before the follow-up questions are asked.

After the follow-up is completed, the results of the interview are reviewed and codes are entered into the system by the matching clerks. An outlier review is also conducted for clusters with high weighted nonmatch and erroneous enumeration rates. "Journals" will be written for these outlier clusters. There will be documentation for all outlier clusters.

The next step in the process is for the missing data work and the dual system estimation within post-strata. See the references in Section 1.0 for more details.

Attachment: Workflow for the Accuracy and Coverage Evaluation

